BERJ | British Educational Research Journal

BERA

British Educational Research Journal
Vol. 43, No. 2, April 2017, pp. 193–212

DOI: 10.1002/berj.3264

The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'

George Leckie* and Harvey Goldstein
Centre for Multilevel Modelling, Graduate School of Education, University of Bristol, UK

Since 1992, the UK Government has published so-called 'school league tables' summarising the average General Certificate of Secondary Education (GCSE) 'attainment' and 'progress' made by pupils in each state-funded secondary school in England. While the headline measure of school attainment has remained the percentage of pupils achieving five or more good GCSEs, the headline measure of school progress has changed from 'value-added' (2002–2005) to 'contextual value-added' (2006–2010) to 'expected progress' (2011–2015) to 'progress 8' (2016–). This paper charts this evolution with a critical eye. First, we describe the headline measures of school progress. Second, we question the Government's justifications for scrapping contextual value-added. Third, we argue that the current expected progress measure suffers from fundamental design flaws. Fourth, we examine the stability of school rankings across contextual value-added and expected progress. Fifth, we discuss the extent to which progress 8 will address the weaknesses of expected progress. We conclude that all these progress measures and school league tables more generally should be viewed with far more scepticism and interpreted far more cautiously than they have often been to date.

Keywords: contextual value-added; expected progress; progress 8; school league tables

Introduction

In England, so-called 'school league tables' summarising the average educational performances made by pupils in each state-funded secondary school have been published annually since 1992 (DfE, 2016a). These tables, derived from pupils' General Certificate of Secondary Education (GCSE) examination results, form a fundamental component of the Government's school accountability by results regime. The tables have their origins in the 1980, 1988 and 1992 Education Reform Acts, which introduced the national curriculum, high-stakes testing and market forces to the education system. This legislation received support from a number of senior academics involved with the Government's Task Group on Assessment and Testing (TGAT) (Black, 1988). TGAT argued both for a national curriculum and the publication of 'unadjusted' test and examination results school by school.

Schools' performances in these tables inform the inspections carried out by the Office for Standards in Education (Ofsted—the school inspectorate system). Schools

*Corresponding author. Centre for Multilevel Modelling, Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol BS8 1JA, UK. E-mail: g.leckie@bristol.ac.uk.

judged underperforming face various sanctions, including increased scrutiny, potential takeover by neighbouring schools and even closure. The tables also play a role in facilitating the quasi-market in education by informing parental school choice. This policy context has been well documented (e.g. West & Pennell, 2000; West, 2010). Our focus in this paper is on statistically critiquing the different headline measures of school 'attainment' and 'progress' which have featured in these tables and played a central role in holding schools to account over the last 25 years.

School *attainment* measures aim to report the average *status* of pupils *at the end* of secondary schooling (year-group 11, age 15/16). The Government's headline measure of school attainment has always been the percentage of pupils achieving five or more GCSEs (or equivalent qualifications judged to be of a similar difficulty; DfE, 2015b) at grade A* to C (5 A*–C; the A* grade was introduced in 1994 to differentiate between top and lower A grades), and since 2006, two of these GCSEs have had to include English and mathematics. For those less familiar with the English education system, see Section S1 of the supplementary materials where we provide an overview together with a summary table (Table S1) of the headline attainment and progress measures. Nationally, 57% of pupils achieved 5 A*–C in 2014 (DfE, 2015c). Unfortunately, this measure is frequently misinterpreted as a measure of the quality of schools. For example, if one school's 5 A*–C percentage exceeds another school's percentage, that difference is all too often attributed solely to a supposed difference in the educational effectiveness of the two schools. However, such an interpretation is invalid, as 5 A*–C confounds any true effect a school has with the composition of each school's intake: schools with higher-attaining pupils at intake will tend to score higher at GCSE, irrespective of the effectiveness of schooling provided. Such straightforward comparisons of average test or examination results are often referred to as 'unadjusted', since no attempt has been made to allow or adjust for such possible confounding effects. Another long-standing criticism of 5 A*–C is that it perversely incentivises schools to concentrate their efforts and resources on pupils at the GCSE grade C/D borderline (West & Pennell, 2000; NAO, 2003; Wilson *et al.*, 2006; West, 2010). Other unintended consequences of the high-stakes nature of 5 A*–C include 'teaching to the test' at the expense of teaching a broader curriculum (Goldstein, 2004), and the practice of entering pupils for 'easier' qualifications (Wilson *et al.*, 2006) and examinations multiple times (Taylor, 2016). Concerns have also been raised about increased anxiety and stress among schools and pupils, as well as pressures on oversubscribed schools to 'cream skim' pupils who are likely to do well on these tests and select out those likely to do poorly.

School *progress* measures aim to report the average *growth* made by pupils *across all five years* of secondary schooling (ages 11 to 16). Progress measures are widely considered the fairer and more meaningful way to compare the effectiveness of schools, for both school choice and accountability purposes, as they implicitly attempt to adjust for what are often substantial differences in the composition of pupils' prior attainments and other characteristics between schools at intake. Our focus in this paper is therefore on school progress measures for school accountability; see Leckie and Goldstein (2009, 2011a) and Wilson and Piebalga (2008) for discussions of issues specific to school choice. In contrast to the headline measure of school attainment, the headline measure of school progress has changed multiple times: from

'value-added' (2002–2005) to 'contextual value-added' (2006–2010) to 'expected progress' (2011–2015) to 'progress 8' (2016–).

The aim of this paper is to explore this evolution of school progress measures with a critical eye. First, we describe the headline measures of school progress. Second, we question the Government's justifications for scrapping contextual value-added (CVA). Third, we argue that the current expected progress (EP) measure suffers from fundamental design flaws. Fourth, we examine the stability of school rankings across CVA and EP. Fifth, we discuss the extent to which progress 8 (P8) will address the weaknesses of EP.

Headline measures of school progress

In 2002 the Government introduced 'national median line' 'value-added' (VA1). VA1 measured how much better (or worse) each school performed in the GCSE examinations than predicted by their pupils' attainments at intake. Specifically, schools' scores were derived as simple school averages of the difference between each pupil's GCSE score in their 'best eight' GCSE and equivalent qualifications and the median score achieved nationally by pupils with the same prior attainment as measured in the end of primary schooling Key Stage 2 (KS2) tests (year-group 6, age 10/11). VA1 was criticised for failing to account for school differences in pupil socioeconomic and demographic characteristics, which had been shown to predict GCSE scores even after adjusting for prior attainment (NAO, 2003; Ray *et al.*, 2009). As such, VA1 was argued, like 5 A*–C, to be biased in favour of schools with more socially advantaged intakes. VA1 was also criticised for failing to communicate the statistical uncertainty surrounding what were in effect the Government's first attempts to estimate the underlying quality or effectiveness of individual institutions.

In 2006 the Government replaced VA1 with CVA, and this measure ran until 2010. CVA attempted to better separate schools' 'true' effects from the composition of their intakes. Conceptually, CVA scores were still school-level averages of the difference between pupils' actual and predicted GCSE scores, but now pupils' predicted scores were calculated as a flexible function of not only their KS2 test scores when they started secondary schooling, but also their age, gender, ethnicity, socioeconomic status [as proxied by free school meal (FSM) eligibility] and various other pupil and school characteristics. These calculations were achieved via fitting a simple multilevel model to the data (Goldstein, 2011) (see Section S2 in the supplementary materials for technical details). The scores were also presented with 95% confidence intervals to communicate the imprecision with which they were estimated.

In 2011 the Government scrapped CVA citing, among other reasons, that it was difficult for the public to understand and that by adjusting for school differences in pupils' socioeconomic and demographic backgrounds, CVA entrenched low educational aspirations in disadvantaged pupil groups (DfE, 2010d). These justifications have gone largely unchallenged in the academic literature. In its place, the Government introduced two new measures of school progress. The first measure, referred to as simply 'value-added' (VA2), simplified the CVA measure by basing pupils' predicted GCSE scores solely on their KS2 scores. Conceptually it was therefore a return to the simplicity of VA1. However, it is the second of the two new progress measures,

EP, which has in effect become the Government's headline measure of progress since 2011.

EP, in contrast to CVA, is not a value-added-based approach. The measure, which is reported separately for English language and mathematics, is calculated as the percentage of pupils in each school who 'make the progress expected of them' during secondary schooling, defined for all pupils as three (or more) national curriculum levels (see Section S1 in the supplementary materials for background information on national curriculum levels). Thus, for example, pupils achieving level 4 in their English KS2 tests (i.e. middle prior attainers) are expected to progress three national curriculum levels to grade C (or higher) in that subject at GCSE; meanwhile pupils achieving level 5 (i.e. high prior attainers) are expected to progress to grade B (or higher). Importantly, the measure does not take into account school differences in pupils' socioeconomic and demographic backgrounds. The measure is also published without confidence intervals. Nationally, 72% of pupils made EP in English in 2014, while 66% made EP in mathematics (DfE, 2015c).

EP also plays a central balancing role in the current minimum levels of performance, or 'floor standards', by which the Government judges schools to be 'underperforming' (DfE, 2010d). A school is judged underperforming if less than 40% of pupils achieve 5 A*–C; however, schools are exempted if their EP scores exceed the national median values in both English and mathematics. Schools judged underperforming face increased scrutiny from Ofsted, potential takeover by neighbouring schools (especially so-called 'academy sponsors'—chains of schools run by charitable or commercial organisations outside the control of their local authorities; HoCL, 2015) and even closure. Nationally, 330 schools (11% of all schools) were judged underperforming in 2014 (DfE, 2015c).

Recently, several education commentators have drawn attention to perceived peculiarities of EP, noting in particular that EP appears biased in favour of high prior attainers (Bostock, 2014; Dracup, 2015; Stewart, 2015). However, we are not aware of any formal studies which examine this and related statistical issues surrounding EP.

Looking to the future, in 2016 the Government is implementing a new school accountability system including new floor standards (DfE, 2016c). As part of this they will scrap EP and introduce a new headline progress measure, P8. P8 marks a return to a value-added-based approach. P8 will be defined in terms of a new measure of GCSE attainment, 'attainment 8' (A8), defined as a pupil's total point score measured across GCSE English and mathematics and six further subjects. The list of approved subjects (DfE, 2016c) is stricter (more academic) than those allowed under 5 A*–C and CVA (and VA1 and VA2). Schools' P8 scores are then simple averages of the differences between pupils' A8 scores and the national average A8 scores of pupils with the same prior attainment. Like EP (and VA1 and VA2), but in contrast to CVA, P8 will make no adjustments for pupil socioeconomic or demographic characteristics. In contrast to EP (and VA1) it will, however, report statistical uncertainty via 95% confidence intervals. P8 will also replace EP in the Government's floor standards. A school will now be judged underperforming if its pupils score on average half a grade lower than predicted and if this difference is statistically significant.

Why was CVA scrapped?

The Government's 2010 Schools White Paper lays out its reasons for withdrawing CVA (DfE, 2010d, p. 68):

> We will put an end to the current 'contextual value added' (CVA) measure. This measure attempts to quantify how well a school does with its pupil population compared to pupils with similar characteristics nationally. However, the measure is difficult for the public to understand, and recent research shows it to be a less strong predictor of success than raw attainment measures. It also has the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances, which we think is wrong in principle.

In this section we examine these three justifications in turn.

Hard to understand

There is certainly merit in their first justification, namely that CVA was hard for the public to understand. After all, CVA scores were derived from a statistical model which included a large set of covariates and their interactions. However, there was no requirement to understand the technical details of the model in order to interpret the CVA scores, only the general principle of adjusting schools' GCSE examination results for differences in prior attainment and related factors between schools at the start of secondary schooling. Perhaps the real problem was in the way CVA was presented to the public. Table 1, which focuses for simplicity on a single local authority, Bristol, reports CVA scores as well as various other performance measures for schools in 2010, the last year CVA was published. CVA scores measured schools' performances relative to the national average (standardised to have a score of 1000) and therefore had no immediate absolute interpretation; a school's score for one year was not directly comparable with their score the year before. More fundamentally, it was not clear to the public what the CVA unit of measurement was; one had to delve deep into the technical documentation (DfE, 2010c) to find out (a 6-point increase in CVA corresponded to pupils, on average, achieving one grade higher in their best eight GCSEs). The CVA 95% confidence intervals were also largely ignored by the media, and no doubt by the public more generally (Leckie & Goldstein, 2011b). In terms of the latter, the Government might have had more success had it tried to communicate the statistical uncertainty in CVA visually rather than in tabular form, as discussed recently by Leckie *et al.* (2016).

It is interesting to note that the Government has somewhat undermined its 'hard to understand' argument by continuing to apply the methodology which underlay CVA in the VA2 (2011–2015) measure (which only adjusts for prior attainment) (DfE, 2011), although VA2 admittedly has a much lower public profile than CVA ever did. It is also worth noting that the methodology which underlay CVA is effectively the same as that used for Hong Kong's public school performance tables today (SVAIS, 2015). It is also simpler than that underlying many other school performance measures published around the world, such as the Tennessee value-added assessment system (TVAAS, 2015).

Table 1. City of Bristol school league table

School	*n*	5 A*–C	CVA	CVA lower	CVA upper	EP English	EP maths
Ashton Park School	180	49	1003	994	1013	66	70
Bedminster Down School	191	40	989	979	998	74	48
Bridge Learning Campus—Secondary	145	34	1003	993	1014	64	44
Brislington Enterprise College	216	37	970	962	979	60	40
Bristol Brunel Academy	158	45	1005	994	1016	69	62
Bristol Cathedral Choir School	75	75	1002	987	1017	95	77
Bristol Metropolitan Academy	127	39	1011	999	1023	76	61
The City Academy Bristol	183	36	1036	1027	1046	71	49
Colston's Girls' School	68	91	1010	992	1027	100	90
Cotham School	180	77	1016	1006	1026	86	85
Fairfield High School	194	49	1004	994	1014	73	63
Henbury School	161	39	1001	991	1011	66	54
Merchants' Academy	124	25	1010	998	1021	56	26
Oasis Academy Brightstowe	93	29	1028	1015	1041	62	37
Oasis Academy Bristol	115	29	1007	995	1019	56	36
Orchard School	172	37	1005	995	1015	69	51
St Bede's Catholic College	185	72	1006	996	1016	80	71
St Bernadette Catholic Secondary School	152	37	980	969	990	66	47
St Mary Redcliffe and Temple School	207	70	1013	1004	1022	86	75

Notes: *n* = number of pupils at the end of GCSE; 5 A*–C = percentage of pupils with five or more GCSEs (or equivalent qualifications) at grade A* to C; CVA = contextual value-added score (national average = 1000); CVA lower = lower limit of CVA 95% confidence interval; CVA upper = upper limit of CVA 95% confidence interval; EP English = percentage of pupils making expected progress in English; EP maths = percentage of pupils making expected progress in mathematics. *Source:* Table reproduced from www.education.gov.uk/schools/performance/archive/schools_10/pdf_10/801.pdf.

A poor predictor of success

It is less clear what the Government means by its second justification: 'recent research shows [CVA] to be a less strong predictor of success than raw attainment measures'. Unfortunately, it does not cite the research referred to. One possible interpretation is that a school's average GCSE performance ('success') is more strongly predicted by their pupils' average KS2 performance ('raw attainment measures') than by their school's CVA score. While this may well be the case, such a result does not in itself mean that CVA is a poor measure of school effectiveness. Indeed, it would more be a reflection of the relatively small influence that schools have on pupil progress (Rasbash *et al.*, 2010) versus the substantial influence of school differences in the composition of pupil prior attainment.

Expected different progress from different pupil groups

The Government's third justification states that '[CVA] also has the effect of expecting different levels of progress from different groups of pupils on the basis of their

ethnic background, or family circumstances. . .' However, CVA did not a priori expect different levels of progress from different pupil groups, rather it adjusted for such differences if they arose. In reality, nationally some pupil groups *do* make less progress than others and this must be adjusted for if we are to make fair comparisons between schools, as otherwise we penalise schools with a disproportionately high number of pupils in these groups. For example, the 2010 CVA model results (DfE, 2010b) show that male pupils, older pupils, pupils with FSM eligibility, pupils in care, pupils with a special educational needs (SEN) statement, mobile pupils and pupils living in deprived neighbourhoods all make less progress than their otherwise equal peers. Pupils who speak English as an additional language and all ethnic minority groups make more progress than White British pupils with the exception of White Irish travellers and White Gypsy/Roman pupils. By adjusting for ethnic background and family circumstances, CVA for the first time rewarded schools for their efforts with harder to teach pupil groups.

The Government expand on its third justification as follows (DfE, 2010d, p. 68):

> It is morally wrong to have an attainment measure which entrenches low aspirations for children because of their background. For example, we do not think it right to expect pupils eligible for free school meals to make less progress from the same starting point as pupils who are not eligible for free school meals (particularly once the introduction of the Pupil Premium ensures that schools receive extra resources for pupils from poorer backgrounds). We should expect every child to succeed and measure schools on how much value they add for all pupils, not rank them on the make-up of their intake.

The Government is arguing that by adjusting for pupil background characteristics, CVA led to a system-level acceptance that socially and other disadvantaged pupil groups *will* make less progress than their more advantaged peers. In other words, it argues that CVA contributed to the lower aspirations and expectations that some schools and teachers hold for their working-class pupils relative to their more advantaged peers.

One mechanism through which this is likely to have occurred is some schools starting to use the published CVA model to set differential GCSE targets for future cohorts of pupils with different socioeconomic and demographic characteristics, even when they had the same prior attainment. Indeed, the DfE gives the following warning, highlighted in red, on the first worksheet of their 2010 CVA 'ready reckoner' (DfE, 2010b):

> . . . because some existing patterns should not become entrenched (for example boys tending to perform less well than girls), this workbook should not be used to determine what students might achieve in the future or in different circumstances.

It goes on to state, in the supporting technical guide: 'CVA should not be used to set lower expectations for any pupil or group of pupils' (DfE, 2010a). However, the CVA model was never intended for this purpose, and that it might have been used in this way reflects the perverse incentives and negative side-effects which so often arise in high-stakes school accountability systems. Whether CVA did entrench low aspirations for poor pupils via this or some other mechanism is not possible to answer using the data at hand. What is known is that the DfE's view is not universally held. In 2013, Brian Lightman, the then General Secretary of the Association of School and

College Leaders (ASCL; the ASCL is a teaching union for secondary school leaders which works to shape national education policy) said the 'ASCL never believed that CVA lowered expectations' (TES, 2013).

Finally, the statement that '[CVA ranks pupils] on the make-up of their intake' suggests a fundamental misunderstanding. CVA explicitly adjusted for as many of the observed differences between schools' intakes as possible, in order to remove their influence from schools' rankings. In contrast, it is when one ignores these differences that one implicitly ranks schools on the make-up of their intake. The aim of CVA is to adjust as fully as possible for all factors for which measurements are available, driving schools' results which can be considered beyond the control of the school, the most important of which are school differences in student composition. To accept this argument in the case of prior attainment, as the Government clearly does, but not for other background characteristics suggests a misunderstanding of the nature of a measure that is to be used as an accountability instrument.

Statistical flaws with EP

The Government's introduction of EP can be seen as an explicit attempt to address the perceived flaws in CVA by providing a school progress measure which is both easier for the public to understand and which is blind to all differences between schools' intakes other than prior attainment. EP, however, suffers from a number of its own fundamental design flaws. In this section we explain and illustrate these flaws using the 2014 school league table data. These data report the EP scores for 3033 mainstream secondary schools whose pupils sat their GCSE examinations in 2014 and their KS2 tests in 2009. To these school-level data we merge the underlying data on pupils' individual EP and KS2 scores from the national pupil database (NPD), the data from which the school league tables are derived (DfE, 2016b).

Borderline effects

EP perversely incentivises schools to concentrate their efforts on pupils who are borderline in terms of making EP. This can be seen by considering Table 2, which shows the GCSE target grade associated with each KS2 level. At the pupil level, EP is a binary measure of progress. Pupils either make their target grade or they do not. There is no middle ground. There is no partial reward for pupils who just miss their target grades, nor any additional reward for pupils who surpass their target grades. The net result is that schools are incentivised to focus their efforts on the subset of pupils at the cusp of making three levels of progress. There is no incentive to work with pupils who are unlikely to make this much progress or to stretch pupils beyond this.

This criticism is essentially the same as that long levelled at 5 A*–C, where schools are known to concentrate resources on C/D borderline pupils. It is therefore unfortunate that EP, developed some 20 years after the introduction of 5 A*–C, should suffer from essentially the same design flaw. Golden *et al.* (2002) and Wilson *et al.* (2006) show that schools engage in a wide range of strategies to support C/D borderline pupils and, given the high-stakes nature of EP, it seems likely that similar strategies

Table 2. Table showing how expected progress in English and mathematics is calculated

	GCSE target grade/level									
	?	U	G	F	E	D	C	B	A	A*
KS2 level	?	2	3	4	5	6	7	8	9	10
?	No	No	?	?	?	?	?	Yes	Yes	Yes
W	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
3	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
4	No	No	No	No	No	No	Yes	Yes	Yes	Yes
5	No	No	No	No	No	No	No	Yes	Yes	Yes

Notes: W = working towards level 1; ? = no result; No = EP not made; Yes = EP made.
Source: Table reproduced from DfE (2015a).

are being used for EP borderline pupils. These include assigning borderline pupils to separate classes, mentoring, homework clubs and Saturday revision classes.

Biased in favour of high prior attainers

EP is severely biased in favour of schools with high prior attaining intakes. Figure 1 illustrates this by presenting the national percentage of pupils making EP in English and mathematics in 2014 separately by KS2 level. We restrict the figure to pupils performing at level 3, 4 or 5 (over 95% of all pupils). The percentage of pupils making EP increases substantially with KS2 level, especially in mathematics, suggesting that

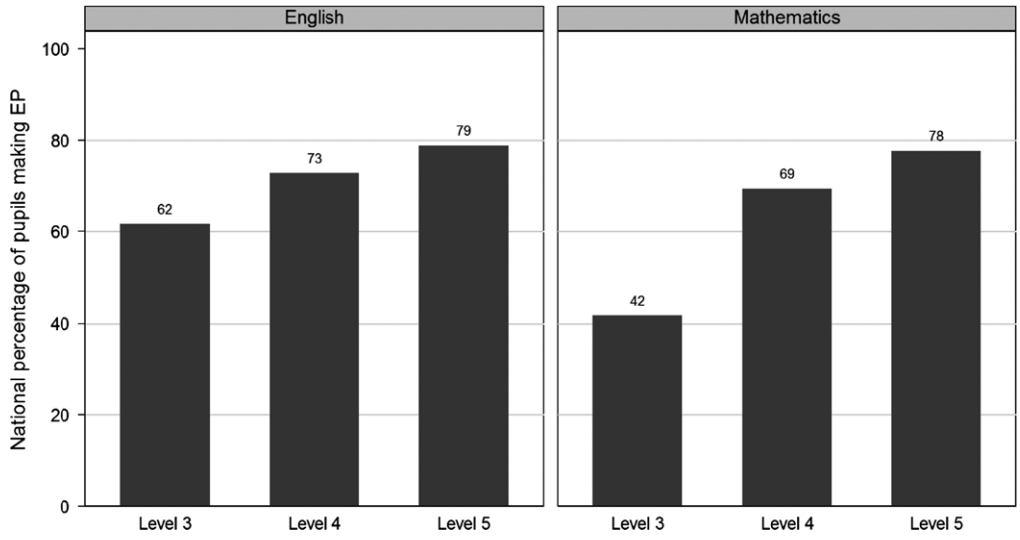

Figure 1. National percentage of pupils making expected progress during secondary schooling against KS2 levels (3, 4 and 5) in 2014, reported separately for English and mathematics.

it is harder for low prior attainers to make three levels of progress than it is for high prior attainers. One reason why low prior attainers struggle more to achieve their target grades might be that they receive lower levels of support than their higher prior attaining peers, but we cannot assess the plausibility of this or other potential explanations from the data at hand.

In terms of school league tables, the implication is that schools' EP scores will very much be driven by their mean intake attainment. Thus, EP, in contrast to CVA, is not a pure progress measure of school performance, but neither is it a pure attainment measure such as 5 A*–C. EP in effect penalises schools with low prior attaining intakes and rewards those with high prior attaining intakes. It follows that, as with 5 A*–C, schools are perversely incentivised to select and subsequently concentrate their efforts on high prior attainers at the expense of their lower prior attaining peers, since the former require less resources to make their target grades (West & Pennell, 2000). To what extent schools have been influenced by these incentives is hard to say, but their existence at all is cause for concern.

Might this 'design flaw' have been in some sense intentional? After all, an argument could be made that EP deliberately sets especially aspirational expectations for low prior attaining pupils vis-à-vis their high prior attaining peers in order to bring about a system-wide narrowing of the attainment distribution. However, if this were the case one might expect a more realistic, tailored and achievable setting of aspirational target grades for low prior attainers than that implied by the edict that all pupils should make three levels of progress irrespective of their starting attainment. It is helpful at this point to view EP from the perspective of a 'categorical', 'transition' or 'transition matrix' model of attainment growth (Castellano & Ho, 2013). In the small literature on this class of model, Table 2 is then referred to as a 'value table', and the associated school performance measure is calculated as the average transition value across the pupils in each school (Hill *et al.*, 2006). In the current case, the transition values are binary (0 = EP not made; 1 = EP made). However, it is perfectly possible and preferable to assign a range of transition values in order to far more intelligently incentivise schools to concentrate their efforts on pupils at particular points in the prior attainment distribution. This would involve policymakers, ideally in conjunction with a wider panel of experts and stakeholders, first making careful value judgements as to the relative merit of each possible transition and then increasing or decreasing the transition values to communicate to schools where they should concentrate their resources (Hill *et al.*, 2006).

While Figure 1 suggests that the probability of making EP increases monotonically with prior attainment, Figure 2 shows the true relationship to be more complex. Figure 2 presents a bar chart of the national percentage of pupils making EP against KS2 sub-levels, plotted separately for English and mathematics. There are three KS2 sub-levels within each KS2 level, and so KS2 sub-levels provide a more finely graded measure of prior attainment than KS2 levels. (See Section S1 in the supplementary materials for further details and Figure S1 for an equivalent plot in terms of pupils' underlying KS2 scores.) The figure reveals a sawtooth (zigzag) relationship between EP and prior attainment, whereby the national percentage of pupils making EP no longer increases monotonically with prior attainment as it did in Figure 1, but now

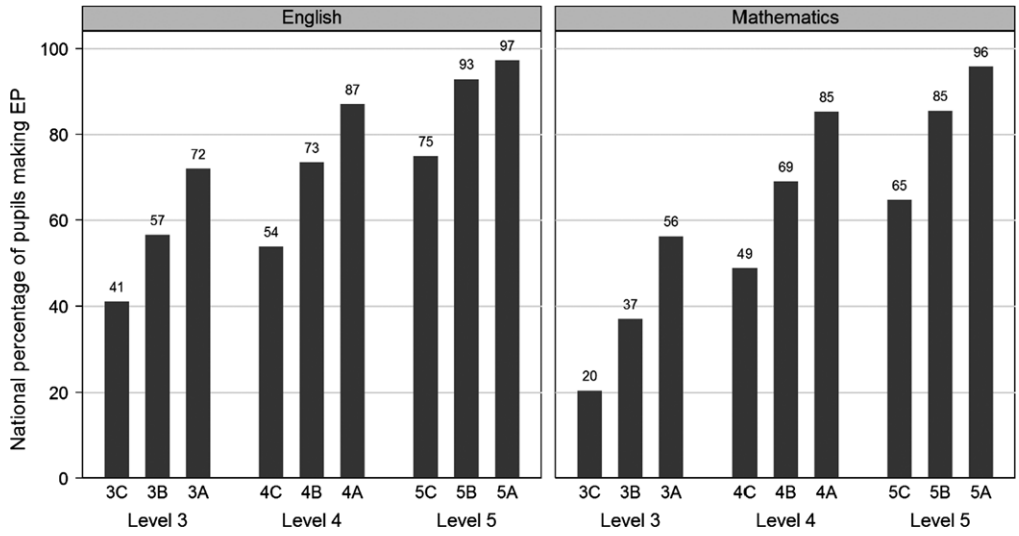

Figure 2. National percentage of pupils making expected progress during secondary schooling against KS2 sub-levels in 2014, reported separately for English and mathematics.

drops dramatically as we move from the top of each KS2 level to the bottom of the next. Figure 2 shows, for example, that while 72% of pupils at KS2 sub-level 3A make EP in English, only 54% of pupils at KS2 sub-level 4C do so. The cause of this discontinuity is that while the KS2 sub-level 3A pupils are set a grade D GCSE target, the KS2 sub-level 4C pupils are set a tougher grade C GCSE target (see Table 2). We see a corresponding discontinuity between KS2 sub-level 4A and sub-level 5C, the point at which the GCSE target grade is raised from a C to a B, respectively. Thus, at the thresholds between KS2 levels, EP results in pupils with effectively identical prior attainment being set different target grades.

In terms of the dramatic increase in the percentage of pupils making EP with respect to KS2 score *within* each KS2 level, this is less surprising when one realises that the small number of KS2 levels necessitates a very large number of pupils and consequently a very wide range of prior attainment within each level. Indeed, half of all pupils achieve KS2 level 4 in English and mathematics (53% and 46%, respectively). Thus, even within each KS2 level, schools are perversely incentivised to concentrate their efforts on their higher prior attaining pupils.

However, perhaps the starkest result of all is in mathematics, where 96% of pupils at KS2 sub-level 5A make expected progress, while the corresponding figure for pupils at KS2 sub-level 3C is just 20%. Clearly, setting the same target of three levels of progress for all pupils makes very little sense. In Figures S2 and S3 in the supplementary materials we plot the actual average number of levels of progress made nationally against KS2 levels and sub-levels, respectively. These figures reveal that average progress differs massively by starting attainment. Taking the same example as before, pupils at KS2 sub-level 3C in mathematics make, on average, only 1.1 levels of progress during secondary schooling, while pupils at KS2 sub-level 5A make, on average, 4.2 levels of progress.

Other pupil characteristics

EP takes no account of pupils' socioeconomic and demographic characteristics. In this sense EP does not expect a priori that disadvantaged pupils will make less progress than their more advantaged peers. However, as we have already argued, the reality is that they do, and markedly so. Thus, for any given level of prior attainment, EP will be biased in favour of schools which serve more advantaged pupil groups.

Statistical uncertainty

EP makes no attempt to quantify and communicate the statistical uncertainty in measuring school progress. A simple example illustrates the severity of the problem. Consider a school with 180 pupils, where 70% make EP. (Such a school corresponds to the national average school, both in terms of school size and EP.) The associated 95% Wald binomial confidence interval ranges from 63% to 77%, and so the school has a ± 7 percentage point margin of error. It is interesting to note that a margin of error of this magnitude would be completely unacceptable in any survey or poll of public opinion (YouGov, 2011), but in the current context this uncertainty is completely ignored; the Government publishes no confidence intervals or margins of error for EP. There is therefore no obvious way for users to establish whether measured differences between schools, or differences from national averages and floor standards, are meaningful differences, or whether they more likely reflect chance variation. Users are implicitly encouraged to view EP scores as error-free, potentially damaging the quality of decision making (Goldstein & Spiegelhalter, 1996; Leckie & Goldstein, 2011b). Figure 3 reveals just how serious a design flaw this is by plotting schools' EP

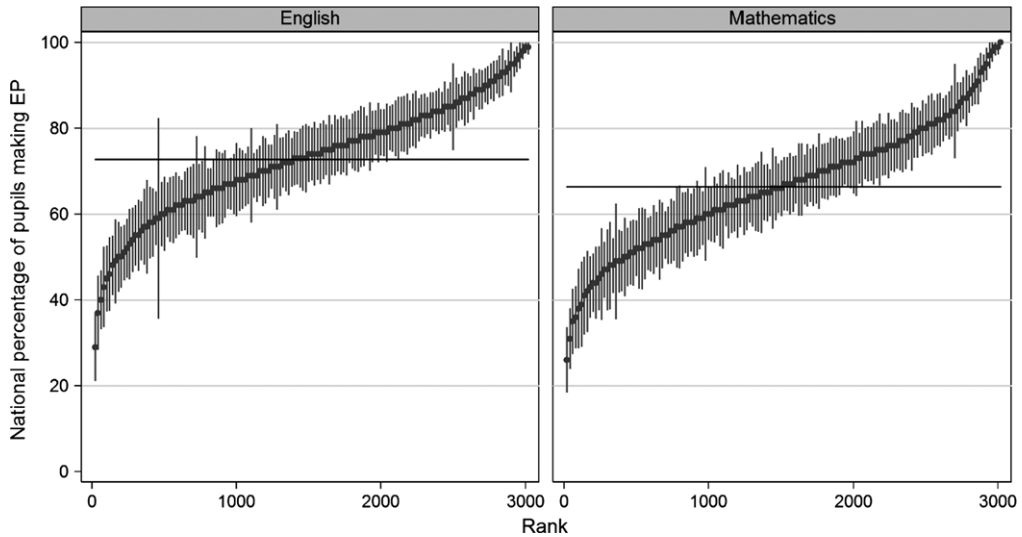

Figure 3. Expected progress scores in 2014 with 95% Wald binomial confidence intervals presented in rank order of magnitude, reported separately for English and mathematics. Higher ranks denote higher performances. The horizontal lines denote the national average EP scores. The confidence intervals are approximate, hence their upper bounds exceeding a value of 100 for a minority of schools with exceptionally high EP scores. For clarity, the plot shows every 20th school.

scores with 95% Wald binomial confidence intervals against their school league table ranking on this measure for all schools in the country. The horizontal lines denote the 2014 national average EP scores of 72% and 66% for English and mathematics, respectively. The figure shows that over a third of schools (38% in English and 36% in mathematics) cannot be statistically distinguished from the national average. Thus, the EP measures are very noisy summary statistics of school progress and to avoid misleading users, this uncertainty must be communicated.

Does choice of school progress measure make a difference in practice?

We have explained how CVA and EP are fundamentally different measures of school progress, both in terms of how they are calculated and in the interpretations they afford. However, if the two measures are highly correlated and lead to similar rankings, then the points we have made could be argued to be largely academic. In this section we therefore show that this is not the case by analysing the 2010 school league table data, the last year for which both CVA and EP appeared concurrently. The data report the CVA and EP scores for 3056 mainstream secondary schools whose pupils sat their GCSE examinations in 2010 and their KS2 tests in 2005.

Correlations between different school progress measures

Table 3 presents Pearson (and Spearman rank) correlations between CVA and EP in English and mathematics in 2010. The table also includes correlations between these progress measures and the Government's headline attainment measure 5 A*–C as well as schools' mean KS2 attainment (averaging across English and mathematics). The correlations between EP and CVA for English and mathematics are low at 0.36 and 0.29, respectively and so ranking schools on the basis of CVA and EP does lead to very different results. Many schools which are ranked high on EP are ranked low on CVA and vice versa. (This is starkly illustrated in Figure S4 in the supplementary materials, a scatterplot of schools' CVA ranks against their EP ranks.) Thus, the additional adjustments that CVA makes for school differences in pupil prior attainment (and socioeconomic and demographic characteristics) over those made by EP are substantial and lead to very different rankings. This is supported by the correlation of just −0.02 between the CVA and KS2 average point score (APS) compared to correlations of 0.64 and 0.67 between EP in English and mathematics and KS2 APS, respectively. EP very clearly inadequately adjusts for school differences in intake attainment. In fact, the prior attainment adjustments that the EP measures make are so slight that EP in both English and mathematics is much more strongly correlated with 5 A*–C than CVA, showing correlations of 0.85 and 0.89 for the two subjects, respectively. Thus, EP appears much closer to being a pure *attainment* measure of school performance than a pure *progress* measure.

The relationship between school progress measures and school mean prior attainment

To investigate further, Figure 4 plots the difference in national ranking between CVA and EP in 2010 against schools' mean KS2 scores. Schools with positive rank

Table 3. Pearson (below the main diagonal of 1s) and Spearman rank (above the main diagonal of 1s) correlation matrices for 5 A*–C, CVA, and EP in English and mathematics in 2010

	5 A*–C	CVA	EP English	EP maths	KS2 APS
5 A*–C	1	0.25	0.84	0.88	0.84
CVA	0.24	1	0.37	0.29	−0.02
EP English	0.85	0.36	1	0.75	0.61
EP maths	0.89	0.29	0.77	1	0.64
KS2 APS	0.87	−0.02	0.64	0.67	1

Notes: Number of schools = 3056. 5 A*–C = percentage of pupils with five or more GCSEs (or equivalent qualifications) at grade A* to C; CVA = contextual value-added score; EP English = percentage of pupils making expected progress in English; EP maths = percentage of pupils making expected progress in mathematics; KS2 APS = KS2 average point score.

differences do better under CVA than under EP and vice versa. The figure shows a strong negative association in each subject, consistent with EP being strongly biased in favour of schools with high prior attaining intakes. We note that the schools with the highest prior attaining intakes, and therefore benefitting most from the design flaws of EP, are 'grammar' schools which set academic entrance exams and therefore select on prior attainment. As well as recruiting the highest prior attaining intakes, we note that grammar schools admit very few FSM and SEN pupils relative to the average school (HoCL, 2016a,b). Within grammar school areas, so-called 'secondary modern' schools take the remaining pupils and therefore tend to have the lowest prior attaining intakes. Thus, secondary modern schools appear most disadvantaged by EP.

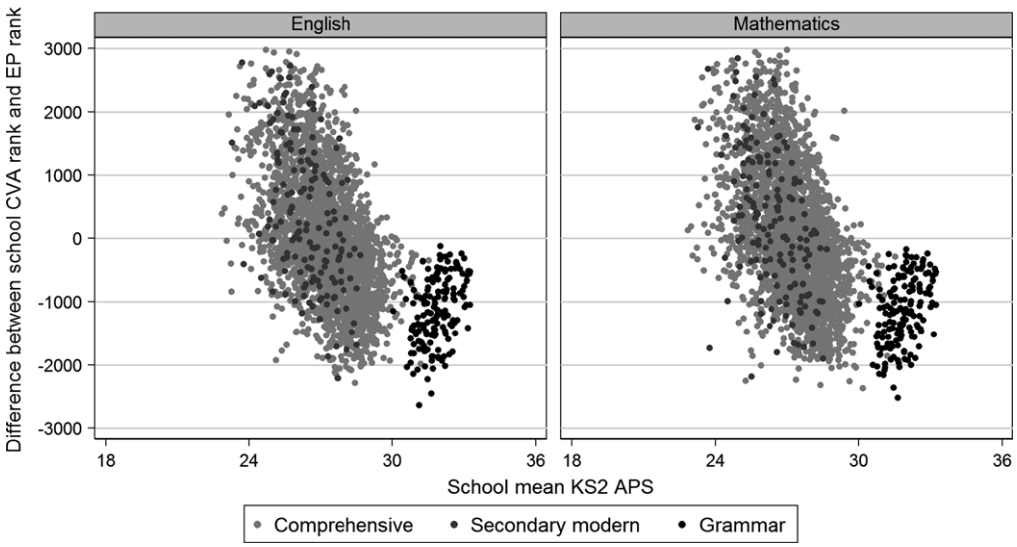

Figure 4. Difference between school CVA and EP ranks against school mean KS2 APS, based on 2010 school league table data, reported separately for English and mathematics. KS2 levels map onto the KS2 point score scale as follows: [18,24] = KS2 level 3 (i.e. low prior attainers); [24,30] = KS2 level 4 (i.e. middle prior attainers); [30,36] = KS2 level 5 (i.e. high prior attainers).

The impact of choice of school progress measure on floor standards

Recall that the Government judges schools to be 'underperforming' if less than 40% of their pupils achieve 5 A*–C but spares schools if, despite this, they exceed the national median in EP in both English and mathematics (DfE, 2015c). Given that just 501 schools (16%) achieved less than 40% 5 A*–C in 2010, one would not expect many of these schools to appear in the top half of schools nationally in EP in English and mathematics. The Government's definition of 'underperforming' is clearly overwhelmingly driven by low attainment on 5 A*–C, and even then the purported 'balancing role' played by the EP measures is undermined by these measures being closer to pure attainment measures than pure progress measures. Indeed, only 37 schools are excused by the high 'progress' they make with their pupils. Another way to compare EP to CVA is therefore to see to what extent these 'underperforming' judgements might be affected were we to excuse schools if they exceed the national median in CVA rather than the national median in EP in both English and mathematics. Whereas we find that 464 schools (15%) are 'underperforming' according to the Government's definition, a substantially lower 313 schools (10%) would be judged underperforming if CVA were used in place of the EP measures. This sizeable reduction in the number of 'underperforming' schools illustrates the far greater balancing role that CVA, a pure progress measure, would have had in contextualising schools' low attainments compared to EP.

Looking ahead to the P8 measure of progress

The forthcoming introduction of P8 into the 2016 school league tables marks a return to a value-added-based approach to measuring school progress. In doing so, P8 will avoid the borderline effects of EP whereby schools are incentivised to focus their efforts on the subset of pupils at the cusp of making three levels of progress (Table 2). P8 should also avoid the systematic bias in favour of high prior attainers (Figure 1) and the illogical sawtooth relationship between progress and prior attainment (Figure 2) exhibited so strongly by EP. P8, in contrast to EP, will be reported with 95% confidence intervals and therefore once again the Government will attempt to communicate the uncertainty in estimating school progress to end users (Figure 3). We note that CVA equally avoided these three design flaws of EP. However, here the similarity between P8 and CVA ends. P8, in contrast to CVA, will continue to make no adjustment for pupils' socioeconomic and demographic characteristics. Presumably this is a continuation of the Government's argument used to justify the withdrawal of CVA, namely that to adjust for such factors would have '. . . the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances . . .' (DfE, 2010d, p. 68). However, as argued above, the choice to adjust or not for such factors is not so simple. Most importantly, by failing to adjust for differences in schools' intakes, P8 will continue to penalise schools serving educationally disadvantaged communities and reward those serving advantaged ones.

In terms of the new floor standards that will also be introduced in 2016 (DfE, 2016c), a school will now be judged underperforming based only on P8. Specifically,

if its pupils score on average half a grade lower than predicted and if this difference is statistically significant. The new floor standards therefore contrast starkly with the existing floor standards, where we have shown that the underperforming status of schools is overwhelmingly driven by low attainment and that progress (in the form of EP) only plays a minor role in balancing these judgements. Thus, this move, and the requirement that schools also be identified as statistically underperforming, should both prove notable improvements on the current floor standards.

Conclusion

In this paper we have discussed the evolution of the headline school progress measure in England from national 'median-line value-added' (VA1, 2002–2005) to 'contextual value-added' (CVA, 2006–2010) to 'expected progress' (EP, 2010–2015) to 'progress 8' (P8, 2016–).

Whereas CVA improved on VA1 by attempting to make as fair comparisons between schools as possible, EP was an explicit ideological shift away from this whereby the Government declared what it wanted to see—three levels of progress in all pupils—and held schools accountable accordingly. P8 represents a shift back to a value-added-based approach in that it once again compares schools to other schools with similar intakes. However, like EP it will remain blind to all socioeconomic differences between schools, beyond those implicit in pupils' prior attainments.

We have critiqued the Government's justifications for scrapping CVA. First, its argument that CVA was hard to understand is compromised by similar and more complex approaches being successfully applied in other schooling systems around the world. Second, the argument that CVA expected different levels of progress from different pupil groups is strongly questionable. CVA recognised that different pupil groups *do* make different progress, and this must be adjusted for in order to make fair comparisons between schools. That schools may have started to use the published CVA models to set differential targets for pupils with the same prior attainment, but different socioeconomic or ethnic status, is a clear misuse of CVA. CVA was not designed for this purpose, and any misuse in this way illustrates the unintended consequences which frequently arise in high-stakes accountability systems.

We presented four fundamental limitations of EP and illustrated these using the 2014 school league table data. First, EP perversely incentivises schools to concentrate their efforts on pupils who are borderline in terms of making EP. Second, EP exhibits an upwards and illogical sawtooth relationship with KS2 score, which severely biases it in favour of schools with high prior attaining intakes. Third, EP ignores the different socioeconomic contexts within which schools operate. Fourth, EP makes no attempt to quantify and communicate statistical uncertainty in measuring school progress.

In terms of our statistical comparison of EP and CVA which used the 2010 school league table data, we find that the two measures are only moderately positively correlated and so the differences in their construction and interpretation are not just academic but lead to fundamental changes in how schools are evaluated. Indeed, EP scores are more strongly correlated with 5 A*–C than with CVA. This suggests that EP is actually closer to being a pure attainment measure of school performance than a pure progress measure. This greatly undermines the 'balancing role' that EP is

purported to play in the Government's 'floor standards' (DfE, 2015c). Indeed, we find that a third of schools judged by the Government to be 'underperforming' in 2010 are in the top half of schools nationally in terms of their CVA performance.

Finally, we described how the introduction of P8 in 2016 marks a return to a value-added-based approach to measuring school progress and in doing so P8, like CVA before it, will address many of the limitations of EP. However, in contrast to CVA, the Government will continue to make no adjustments for school differences in pupils' socioeconomic and demographic backgrounds when they enter their schools, and we think this decision is highly problematic given the substantial impact such differences make on schools' rankings.

There are very likely differential impacts of VA1, CVA, EP and P8 on both schools and their pupils. At the school level those schools with higher prior achieving intakes are likely to appear particularly successful under EP as it only makes a partial adjustment for prior achievement compared to the other measures. Indeed, we have shown that grammar schools' national rankings in 2010 were substantially higher under EP than under CVA. Schools whose intakes are more socioeconomically advantaged are likely to benefit from VA1, EP and P8 compared to CVA, as all of these measures confound this advantage with any true influence of the school. At the pupil level, pupils identified as being borderline in making EP are likely to have benefitted from the move from CVA to EP as schools were suddenly incentivised to focus their efforts on this narrow group. In terms of socioeconomic status, the Government would argue that disadvantaged pupils would benefit under EP (and VA1 and VA2) as their schools would now have to aspire for them to achieve higher than under CVA. However, it could be the case that by judging disadvantaged pupils once again by the same standards as their more advantaged peers, schools may shift their efforts and resources away from harder to teach pupil groups.

We conclude that all these progress measures and school league tables more generally should be viewed with far more scepticism and interpreted far more cautiously than they have often been to date. Our view is that the CVA measure, and more generally the multilevel value-added modelling approach which underlies it, has many advantages over EP and various simpler VA models which have been proposed —including P8. CVA, by virtue of accounting for the richest set of influences on student achievement, is also the progress measure most consistent with the main theoretical models proposed in the educational effectiveness literature (see Reynolds *et al.*, 2014 for a recent review). However, a range of well-documented statistical and more general issues with making quantitative comparisons between schools remain, whatever the measure employed (Goldstein & Spiegelhalter, 1996). A specific statistical issue we have not discussed is differential effectiveness—the notion that schools can have differential effects on different pupil groups—yet this is an important issue when holding schools to account (Nuttall *et al.*, 1989; Goldstein, 1997; Strand, 2016). It is also an issue which the Government has become increasingly interested in, as evidenced by its separate reporting of various headlined attainment and progress measures by pupil subgroups (chiefly with respect to prior attainment and socioeconomic disadvantage).

A more general concern is the degree to which school league tables, progress or otherwise, should be used to hold schools accountable at all. The current

deterministic rule that a school is underperforming if it simultaneously fails floor standards in 5 A*–C and EP, as well as the revised version of this rule when P8 comes into effect, appears overly rigid given the high-stake consequences of such judgements. Many have argued (Willms, 1992; Harris *et al.*, 1995; Goldstein & Spiegelhalter, 1996; Goldstein & Thomas, 1996; Demie, 2003), and we would agree, that school league tables are best used as tools for school self-evaluation and as a first step towards identifying successful school policies and practices. Where they are used by the Government and school inspection systems, they may be better used as monitoring and screening devices to identify schools performing unexpectedly poorly for the purpose of careful and sensitive further investigation (Foley & Goldstein, 2012). Even then, school league tables will be of most use when combined with other sources of school information.

Acknowledgements

This research was funded by UK Economic and Social Research Council grant ES/K000950/1. We are grateful for the helpful comments provided by the Editors and the two reviewers.

References

Black, P. J. (1988) *National curriculum: Task group on assessment and testing – a report* (London, Department of Education and Science and the Welsh Office).

Bostock, M. (2014) *Progress and VA – the test of school quality*. Available online at: mikebostock.com/progress-and-va/ (accessed 01 February 2016).

Castellano, K. E. & Ho, A. D. (2013) *A practitioner's guide to growth models* (Washington, D.C., Council of Chief State School Officers).

Demie, F. (2003) Using value-added data for school self-evaluation: A case study of practice in inner-city schools, *School Leadership & Management*, 23(4), 445–467.

DfE (2010a) *A technical guide to contextual value added (including English & maths) Key Stage 2 to 4 2010 model* (London, Department for Education).

DfE (2010b) *Guide to contextual value added models KS2–4 2010* (London, Department for Education).

DfE (2010c) *Test and examination point scores used in the 2010 school and college performance tables* (London, Department for Education).

DfE (2010d) *The importance of teaching: The Schools White Paper 2010* (London, Department for Education).

DfE (2011) *A guide to value added Key Stage 2 to 4 in 2011 school & college performance tables & RAISEonline* (London, Department for Education).

DfE (2015a) *Key Stage 2 to Key Stage 4 progress measures 2014* (London, Department for Education).

DfE (2015b) *Key Stage 4 performance tables: Inclusion of 14–16 non-GCSE qualifications* (London, Department for Education).

DfE (2015c) *Revised GCSE and equivalent results in England, 2013 to 2014*. Statistical First Release, SFR 02/2015 (London, Department for Education).

DfE (2016b) *Compare school and college performance* (London, Department for Education).

DfE (2016c) *National pupil database* (London, Department for Education).

DfE (2016a) *Progress 8 measure in 2016, 2017 and 2018: Guide for maintained secondary schools, academies and free schools* (London, Department for Education).

Dracup, T. (2015) *2014 Primary and secondary transition matrices: High attainers' performance*. Available online at: giftedphoenix.wordpress.com/2015/01/07/2014-primary-and-secondary-transition-matrices-high-attainers-performance/ (accessed 7 January 2015).

Foley, B. & Goldstein, H. (2012) *Measuring success: League tables in the public sector* (London, British Academy).

Golden, S., Knight, S., O'Donnell, L., Smith, P. & Sims, D. (2002) *Learning mentors' strand study: Excellence in cities*. Report 16/2002 (Slough, National Foundation for Educational Research).

Goldstein, H. (1997) Methods in school effectiveness research, *School Effectiveness and School Improvement*, 8(4), 369–395.

Goldstein, H. (2004) Education for all: The globalization of learning targets, *Comparative Education*, 40, 7–14.

Goldstein, H. (2011) *Multilevel statistical models* (4th edn) (Chichester, UK, Wiley).

Goldstein, H. & Spiegelhalter, D. J. (1996) League tables and their limitations: Statistical issues in comparisons of institutional performance, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 159(3), 385–443.

Goldstein, H. & Thomas, S. (1996) Using examination results as indicators of school and college performance, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 159, 149–163.

Harris, A., Jamieson, I. & Russ, J. (1995) A study of 'effective' departments in secondary schools, *School Organisation*, 15(3), 283–299.

Hill, R., Gong, B., Marion, S., DePascale, C., Dunn, J. & Simpson, M. A. (2006) Using value tables to explicitly value growth, in: R. Lissitz (Ed) *Longitudinal and value added models of student performance* (Maple Grove, MN, JAM Press), 255–290.

HoCL (2015) *Free schools and academies: FAQ*. Number 07059 (London, House of Commons Library).

HoCL (2016a) *Grammar schools in England*. Briefing paper. Number 7070 (London, House of Commons Library).

HoCL (2016b) *Grammar school statistics*. Briefing paper. Number 1398 (London, House of Commons Library).

Leckie, G. & Goldstein, H. (2009) The limitations of using school league tables to inform school choice, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 172, 835–851.

Leckie, G. & Goldstein, H. (2011a) A note on 'The limitations of school league tables to inform school choice', *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 174, 833–836.

Leckie, G. & Goldstein, H. (2011b) Understanding uncertainty in school league tables, *Fiscal Studies*, 32, 207–224.

Leckie, G., Charlton, C. & Goldstein, H. (2016) *Communicating uncertainty in school value-added league tables* (Bristol, Centre for Multilevel Modelling, University of Bristol).

NAO (2003) *Making a difference: Performance of maintained secondary schools in England* (London, National Audit Office).

Nuttall, D. L., Goldstein, H., Prosser, R. & Rasbash, J. (1989) Differential school effectiveness, *International Journal of Educational Research*, 13(7), 769–776.

Rasbash, J., Leckie, G., Pillinger, R. & Jenkins, J. (2010) Children's educational progress: Partitioning family, school and area effects, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 173(3), 657–682.

Ray, A., McCormack, T. & Evans, H. (2009) Value added in English schools, *Education*, 4, 415–438.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C. & Stringfield, S. (2014) Educational effectiveness research (EER): A state-of-the-art review, *School Effectiveness and School Improvement*, 25(2), 197–230.

Stewart, H. (2015) *How to use data badly: Levels of progress*. Available online at: www.localschoolsnetwork.org.uk/2015/03/how-to-use-data-badly-levels-of-progress (accessed 29 March 2015).

Strand, S. (2016) Do some schools narrow the gap? Differential school effectiveness revisited, *Review of Education*, 4(2), 107–144.
SVAIS (2015) *Schools value-added information system technical manual* (Hong Kong, Education Bureau Quality Assurance & School-Based Support Division).
Taylor, R. C. (2016) The effects of accountability measures in English secondary schools: Early and multiple entry to GCSE mathematics assessments, *Oxford Review of Education*, 1–17.
TES (2013) Where you come from matters after all, says Gove, *Times Educational Supplement*, No. 5027, p. 18.
TVAAS (2015) *Tennessee value-added assessment system*. Available online at: www.tn.gov/education/topic/tvaas (accessed 01 February 2016)
West, A. (2010) High stakes testing, accountability, incentives and consequences in English schools, *Policy and Politics*, 38, 23–39.
West, A. & Pennell, H. (2000) Publishing school examination results in England: Incentives and consequences, *Educational Studies*, 26, 423–436.
Willms, J. D. (1992) *Monitoring school performance: A guide for educators* (Oxford, Routledge).
Wilson, D. & Piebalga, A. (2008) Performance measures, ranking and parental choice: An analysis of the English school league tables, *International Public Management Journal*, 11, 344–366.
Wilson, D., Croxson, B. & Atkinson, A. (2006) What gets measured gets done: Headteachers' responses to the English secondary school performance management system, *Policy Studies*, 27, 153–171.
YouGov (2011) *Understanding margin of error*. Available online at: yougov.co.uk/news/2011/11/21/understanding-margin-error/ (accessed 21 November 2011).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Section S1. Background on English education system and national testing.
Section S2. CVA.
Figure S1. National percentage of pupils making expected progress during secondary schooling against KS2 APS in 2014, reported separately for English and mathematics. The magnitude of the hollow circles is proportional to the national number of pupils with that KS2 APS. The dashed vertical lines denote the KS2 level thresholds. Level W = working towards level 1. For clarity, the plot is restricted to values of KS2 APS for which there were at least 100 pupils nationally.
Figure S2. National mean number of levels of progress during secondary schooling against KS2 levels in 2014, reported separately for English and mathematics.
Figure S3. National mean number of levels of progress during secondary schooling against KS2 sub-levels in 2014, reported separately for English and mathematics.
Figure S4. Scatterplot of school CVA ranks against EP ranks, based on 2010 school league table data, reported separately for English and mathematics. Higher ranks denote higher performances.
Table S1. Headline attainment and progress measures since 1992.